Check for updates

Perspective

Systematic online living evidence summaries: emerging tools to accelerate evidence synthesis

Kaitlyn Hair[1], Ⓘ **Emma Wilson[1], Charis Wong[2,3], Anthony Tsang[4],** Ⓘ **Malcolm Macleod[1]** and
Ⓘ **Alexandra Bannach-Brown[5]**

[1]Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, U.K.; [2]Anne Rowling Regenerative Neurology Clinic, University of Edinburgh, Edinburgh, U.K.; [3]Euan Macdonald Centre for Motor Neuron Disease Research, University of Edinburgh, Edinburgh, U.K.; [4]King's Technology Evaluation Centre, King's College London, U.K.; [5]Charité Universitaetsmedizin Berlin, Berlin Institute of Health – QUEST Center, Berlin, Germany

Correspondence: Alexandra Bannach-Brown (alexandra.bannach-brown@charite.de)

OPEN ACCESS

Systematic reviews and meta-analysis are the cornerstones of evidence-based decision making and priority setting. However, traditional systematic reviews are time and labour intensive, limiting their feasibility to comprehensively evaluate the latest evidence in research-intensive areas. Recent developments in automation, machine learning and systematic review technologies have enabled efficiency gains. Building upon these advances, we developed **S**ystematic **O**nline **L**iving **E**vidence **S**ummaries (SOLES) to accelerate evidence synthesis. In this approach, we integrate automated processes to continuously gather, synthesise and summarise all existing evidence from a research domain, and report the resulting current curated content as interrogatable databases via interactive web applications. SOLES can benefit various stakeholders by (i) providing a systematic overview of current evidence to identify knowledge gaps, (ii) providing an accelerated starting point for a more detailed systematic review, and (iii) facilitating collaboration and coordination in evidence synthesis.

The need for systematic review in biomedicine

For hundreds of years, researchers have recognised the need to effectively synthesise research evidence to facilitate evidence-based decision making [1]. Systematic review (SR), a research methodology to locate, select, evaluate and synthesise all evidence on a specific research question based on pre-specified eligibility criteria, emerged in the latter half of the twentieth century as a tool to enable clinicians to make evidence-based decisions about patient care. Evidence from SRs is widely regarded as the highest level of evidence in the evidence hierarchy and is often used to inform improvements in clinical trial design and health care [2]. Over the past decade, there have been growing attempts to apply similar methodologies to summarise and evaluate findings from animal (*in vivo*) and, more recently, *in vitro* laboratory research [3]. Many challenges – including the internal and external validity of preclinical research – may impede the translation of findings from animal and cell-based models to human patients. Preclinical SRs can help expose the reasons behind this discrepancy [4]. For example, poor methodological quality and publication bias in stroke research may overestimate therapeutic benefits observed in animal models [5,6]. SRs of including both clinical and preclinical literature can be helpful to inform selection and prioritisation of candidate drugs in clinical trials, especially in fields with high rates of translational failure such as neurodegenerative diseases [7–9].

Barriers limiting the use of systematic reviews

There are proven benefits of conducting SRs for evidence-based research and decision making. Here, we focus on the application of SRs to preclinical and clinical evidence. However, traditional SR methodology

is no longer optimal for the current research landscape where rates of publication are increasing tremendously [10]. Here, we outline the three key barriers to impact.

Failing to capture all relevant evidence

One of the key strengths of SRs over more traditional narrative literature review is that they aim to capture all the available evidence. However, even SRs are not universally successful in this regard [11–13]. Methodological choices such as inadequate search strategies, narrow research questions and exclusion of grey literature sources (e.g., preprint servers), and factors relating to lack of standardisation in database indexing and delays in indexing, all limit the completeness with which relevant primary research is identified. Systematic searches typically search for specific terms within keywords, title and abstract within bibliographic databases. However, emerging data from ongoing SRs of open field test behaviour and synaptic plasticity in transgenic Alzheimer's models which used full-text searching found that 82% and 69% of relevant manuscripts, respectively, were not identified by searching the keywords, title or abstract [14]. Similarly, in a SR of oxygen glucose deprivation in PC12 cells, human reviewers conducting title and abstract screening erroneously excluded 14% of relevant studies, which were only identified using a broader search coupled with full-text screening [15]. Traditional approaches therefore may not be sensitive enough for preclinical biomedical SRs.

Failing to keep up with the literature

By the time of their publication, SRs are frequently many months or years out of date [16,17]. Manually searching bibliographic databases, screening studies for relevance, and extracting study details is laborious and time-consuming [18]. This issue is amplified in highly research intensive areas, where there could be thousands of potentially relevant studies published every year. For example, to update a review of animal models of neuropathic pain, we performed a systematic search three years after the original and found 11,000 new publications to screen for relevance [19]. Furthermore, updating SRs to incorporate recent evidence is not standard practice [20] and can require just as much effort as conducting the initial review [21,22]. Incorporating new studies into evidence summaries is an ongoing challenge with the rate of publication increasing exponentially each year [8]. Even where updates to SRs are done, they may use different SR methodology [23] making it difficult to contextualise the new findings. For research areas with a high rate of publication, it has been suggested that the median 'survival' of reviews (i.e., period over which they remain up to date) is just under three years [24]. Recent findings suggest that half of Cochrane's reviews were published more than 2 years after the original protocol, and that this gap may be widening over time [25]. It has also recently been reported that 85% of Cochrane's reviews have not been updated for over 5 years [26].

Failing to collaborate and co-ordinate efforts

Without co-ordinated efforts to prioritise and conduct which SRs should be done, the degree of overlap can be significant [27], with little acknowledgement of findings from previous reviews on the same or similar research questions [28]. Collaboration among researchers can help reduce the resource burden required to complete a SR, but this requires researchers to be aware of other reviews being conducted in areas that overlap with their own interests. Greater use of registration tools such as PROSPERO is likely to reduce duplication of effort. More cohesive evidence synthesis communities, with stronger collaboration between primary researchers, information professionals, statisticians and evidence synthesists and key stakeholders and users of evidence synthesis research, would help to address this. Stronger communities could improve how priority questions in biomedicines are identified and addressed, ensure that synthesis research is conducted collaboratively, and increase the likelihood that findings are put into practice in a timely manner [29].

Emerging approaches and 'living' systematic reviews

Setting aside the usefulness of SRs and meta-analyses, alternative approaches such as rapid reviews, or scoping reviews [30] may be sufficient and perhaps more feasible in a given timescale. Evidence gap maps are an emerging concept, where all evidence relating to a broad research question is collected and presented in a user-friendly format, to identify knowledge gaps and inform future research priorities [31,32].

Recently, several groups have sought to facilitate faster evidence synthesis through the application of automation technologies [33]. By incorporating natural language processing and other machine-assisted techniques, the human effort and time required to complete each step of the SR process can be dramatically reduced. Several tools have been developed for performing systematic searches [34,35], record screening [36–38], deduplication [39] and extracting information from publications [40,41]. Through levering such methodologies, the concept of 'living' SRs has been

Table 1 Summary of the features and differences between SOLES, evidence gap maps and living SRs

	SOLES	Evidence gap maps	Living systematic review
Systematic search	Broad domain-specific search terms	Search terms tailored to a broad research area or question.	Narrower search terms, tailored to a specific research question
Eligibility criteria	Broad inclusion criteria including all experiments of relevance to research domain	Inclusion criteria focusing on a particular research question or area, typically include SRs on the topic	Narrower inclusion criteria, focusing on a particular population, intervention and/or outcome
Data annotation	Prioritised annotation of major PICO elements, study quality indicators and other measures (including transparency of reporting)	Prioritised annotation of major PICO elements and study quality indicators	In-depth annotation of all PICO elements, study quality indicators, study design characteristics and other metadata
Meta-analysis	Not currently a feature	In most cases, not appropriate	May be conducted in the context of systematic review if appropriate
Output	An online dashboard with several interactive visual summaries of the latest evidence	A visual overview of evidence (i.e., interventions and outcomes) relevant to research question or area	A detailed synthesis and evaluation of evidence related to a pre-defined, narrow research topic often in the form of a scientific publication
Use cases	Used to gain a high-level overview of a research domain, to identify heavily studied areas or gaps in knowledge, to benchmark quality and transparency measures and to serve as a starting point for more detailed SRs on narrower topics	Used to gain a high-level overview of a research question, to identify heavily studied areas or gaps in knowledge	Used to make decisions regarding future research (evidence-based research), and to inform policy and practice
Updates	Updated continuously in response to emerging evidence (e.g., daily or weekly)	Updated frequently (as described in an individual protocol)	Updated frequently (as described in an individual living systematic review protocol)

Abbreviation: PICO: Population, Intervention, Comparator and Outcome.

proposed as now being a feasible approach [42]. A living SR is a systematically obtained and continually updated summary of a research field, incorporating new findings as and when they emerge. However, due to the heterogeneity of published literature, some human effort will likely always be required to understand and extract detailed information and numerical data from publications.

Systematic Online Living Evidence Summaries (SOLES)

Drawing on the objectives of living SRs and evidence gap maps, and considering the barriers faced by conventional SR approaches, we introduce a new approach for evidence synthesis. **S**ystematic **O**nline **L**iving **E**vidence **S**ummaries (SOLES). SOLES are systematically obtained, continuously updated, curated, and interrogatable datasets of existing evidence, enabled by automation. They are likely of greatest benefit to research-intensive areas with high rates of publication, as seen in much of biomedicine. Here, we describe where SOLES fit into the evidence synthesis ecosystem (see Table 1), describe an exemplar SOLES approach as applied to Alzheimer's disease research (see Table 2), discuss how SOLES may address some of the barriers limiting traditional SRs, and discuss current limitations and future developments of the SOLES approach.

The major differences between SOLES and related approaches lie in the scope of evidence included and the application of automation technologies. Most living SRs have a narrow, clearly defined research question, whereas SOLES aim to summarise much larger areas of research. SOLES and evidence gap maps share many similarities in their overall approach and immediate use-cases, however SOLES encompass broader research domains and aim to present a heterogeneous range of visual summaries, trends, and functionality to research users. A comparison of each approach is shown in Table 1. Importantly, evidence gap maps have largely been applied to collate evidence from clinical studies from the social sciences, where SRs are routinely conducted. As such, evidence from existing SRs are an important feature. SOLES were developed with the initial aim of curating the vast quantity of diverse evidence in preclinical models of human disease to facilitate SRs in this area and drive research improvement.

The basic workflow for a SOLES is illustrated in Figure 1. SOLES begin with a wide search strategy across several biomedical databases and use less stringent criteria for inclusion – for example, including all laboratory experiments in any animal model of Alzheimer's disease. Full texts of included studies are retrieved – for open access publications via Unpaywall, and for others via institutional subscriptions using Crossref [46–50]. High-level data are then extracted to categorise the research – e.g., by disease model(s), intervention(s), and outcome measure(s). Much of this process can be automated, either through machine learning approaches or simple text mining e.g. [51]. In addition, high-level quality measures may also be extracted, with quality of reporting assessed using text-mining or machine
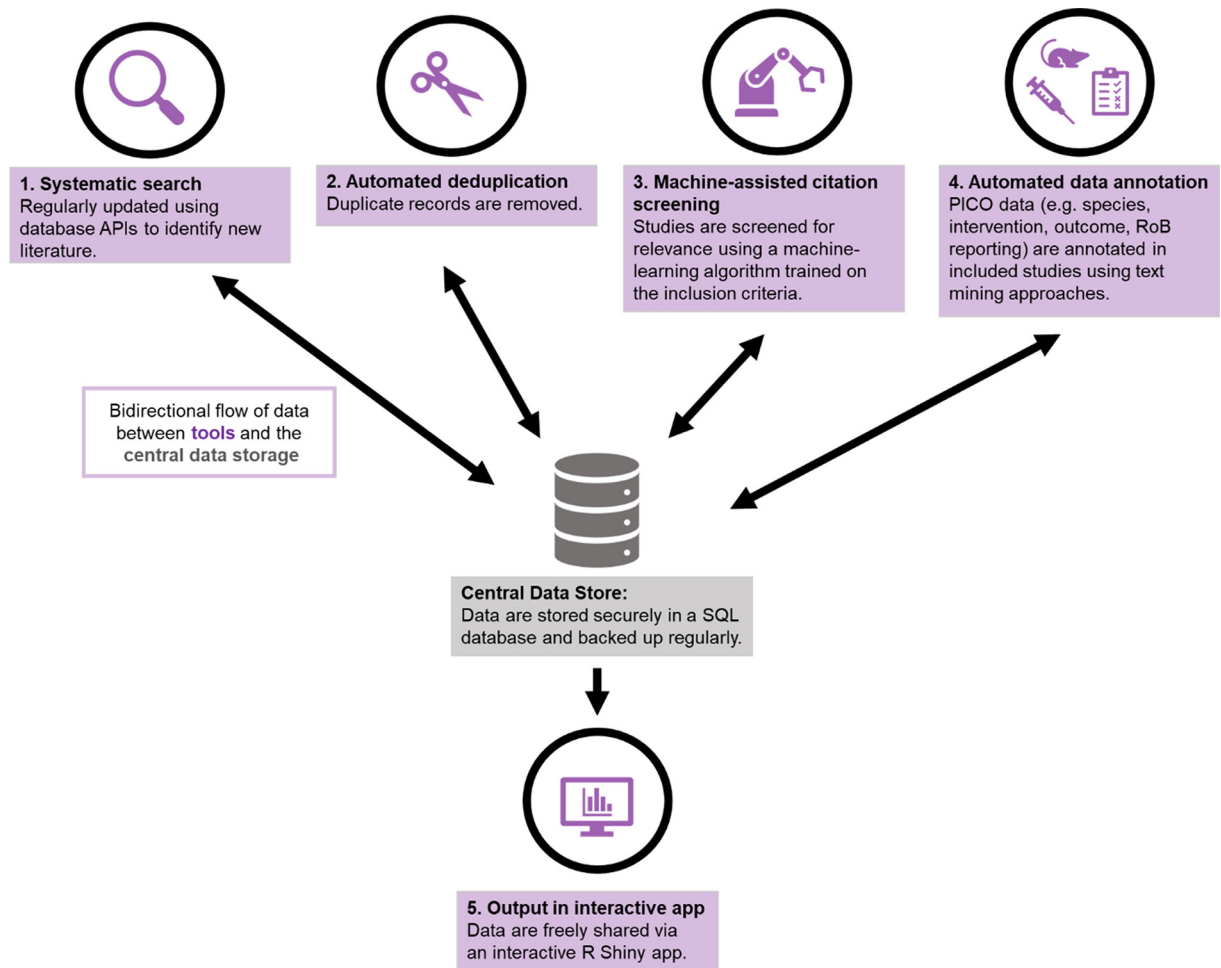
Figure 1. The automated SOLES workflow

learning tools (e.g., [41,43]). This curated dataset and accompanying visualisations can then be made accessible to the research community via an interactive web application, such as R Shiny Apps (RRID:SCR_001626). This enables key stakeholders such as researchers, regulatory agencies, and funders to quickly ascertain the quantity and quality of research evidence, both overall and within each subgroup of studies employing each model, testing a certain treatment, or measuring a specific outcome. Furthermore, if the SOLES dataset sufficiently captures the existing literature, those wishing to perform a SR within that research domain can download a filtered set of references as a starting point, eliminating one of the most time-consuming tasks in the systematic review process [12].

The SOLES approach integrates automated processes to benefit end users. Despite the existence of automation tools, as summarised earlier in this paper, there are significant practical and knowledge barriers which reduce their uptake by individual SR teams [44]. Many tools exist as isolated use-cases which lack integration with other parts of the systematic workflow and may not be immediately available to researchers without sufficient technical expertise [45]. SOLES provides a single platform that integrates several tools in a pipeline to benefit researchers, evidence synthesists, and evidence users and decision makers (e.g., funders and regulatory agencies). There are direct benefits in obtaining a single evidence base on a topic, through a concentrated and collaborated effort, on which further, more detailed, reviews and primary research may be based.

Maximising the potential of research evidence

SOLES also aim to transform the way we make sense of, and evaluate, research evidence. Next, we discuss how the SOLES approach can offer solutions to some of the barriers which limit more traditional SRs and benefit key stakeholders of evidence.

Capturing a wider range of evidence

Developing search strategies for SRs has always been a balancing act between trying to ensure that no relevant evidence is missed (i.e., the search has high sensitivity) without capturing too many irrelevant publications to screen (i.e., the search has high specificity). Using automation approaches for record screening allows for broader systematic search strategies with higher sensitivity, to prioritise sensitivity over specificity. At the most extreme, one might imagine record screening being applied to an entire bibliographic database rather than to the results of a systematic search, the specificity being derived from downstream, automated tasks applied to full text articles. However, current computational limitations and tool performance mean that this is not possible at present.

Further, the sensitivity of the SOLES approach to literature searching does not rely on the presence of specific PICO elements in the title or abstract of publications. The pool of potentially relevant publications is much larger but, as we have shown in exemplar SOLES projects, the evidence can then be further subdivided into PICO categories by applying natural language processing techniques (regular expressions or machine learning algorithms) on the full-text of the publication.

A highly sensitive search to retrieve as many relevant citations as possible comes with the caveat that many individual citations will have been identified from several databases. With extremely large searches, this issue is amplified and difficult to manage using conventional deduplication tools. We have recently developed ASySD – the **A**utomated **Sy**stematic **S**earch **D**eduplication tool [39]. ASySD is an open-source tool created with SOLES functionality in mind, and allows the user to preferentially retain certain citations within a group of duplicate citations. For example, users can elect to retain older citations already present (and potentially annotated) in the dataset, or to retain the record with the most complete information.

Using SOLES as a starting point for obtaining the latest and most relevant research enables researchers and funding agencies to quickly gain a credible and systematic overview of current evidence. This can facilitate decision making, ensuring that future research is focused on the areas of greatest need and minimising research waste.

Keeping up to date with the literature

A major goal of SOLES projects is to synthesise new research evidence as soon as it becomes available, thus preventing any delay in making use of that evidence to inform future research and decision making. SOLES developers can implement fully automated searches for publications using accessible Application Programming Interfaces (APIs) – which *exist* for several platforms including PubMed (NBCI), Scopus, Dimensions and Web of Science collections – and schedule these search queries to run at scheduled time periods. Attempts to automate some searches have been hampered by issues in retrieving citation data of the same quality (with abstracts and other meta-data present) and cost for access. Depending on institutional access and specific database needs, this step may require some human input. The retrieval of new citations on a regular basis also facilitates attempts to update SRs with minimal additional effort.

Systematic review questions generally focus on specific PICO (population, intervention, comparison and outcomes). To enable SOLES users to identify the latest relevant research, categorisations can be added to publications using regular expressions (regex) search dictionaries for animal models, interventions and outcome measures. A regex is a sequence of characters relating to a pattern of text, and are commonly used in library and information science and more widely. Overall, these approaches are time-saving and can be highly sensitive, especially if sufficient effort is placed into designing regex probes to account for common synonyms and deviations in punctuation and spelling (Bahor et al., 2017). In this way, using a curated citation data set derived from a SOLES platform can be an accelerated starting point for a more detailed SR.

To ascertain study quality, and to gain insights over whether primary study quality may be improving over time, similar text-mining tools can be applied to research articles to measure reporting quality or measures to reduce the risk of bias [43]. More recently, machine learning classifiers tailored to preclinical research have been developed to assess risk of bias reporting [41].

The automated PICO annotation and quality assessment tools allow for flexibility of the approach to analysis, depending on the needs of the research user. For example, the quality of a whole domain (e.g., animal research in Alzheimer's disease) of evidence can be quickly determined for funders making decisions in the design of new funding calls. Alternatively, current data on highly specific research questions can be interrogated by primary researchers designing a new experiment. Together, these approaches bring us a few steps closer to making use of new evidence at a much quicker pace. Further, measuring research quality indicators in an automated way may allow for the benchmarking of research improvement activities over time [52,53].

Collaboration and co-ordination

SOLES approaches require substantial human effort to set-up and ensure that the automated processes used to screen and annotate information function optimally. Many of the automation tools used to power SOLES require initial human input. For example, record screening uses supervised machine learning algorithms, requiring in turn large, curated training datasets. A random subset of search results should be screened manually to train an algorithm on the inclusion and exclusion criteria of a particular SOLES project, even if such an algorithm has already been validated in a different context. This calls for efficient collaboration and co-ordination to ensure that training data can be collected quickly, and that data are reliable enough to train the algorithms.

Large projects and collaboration provides an opportunity to use a crowdsourcing approach, which, if well-co-ordinated, allows for quicker data collection. Crowdsourced approaches work best if participants are given 'micro-tasks' to perform, such as screening or annotation of only a few pieces of information at a time [54]. Participating in a crowdsourced project can be beneficial, especially for junior researchers or students who want to get involved in systematic review but do not have the time or skills to conduct a full review themselves. However, it is important to consider whether you have enough resources to train, manage and enthuse your crowdsourced participants during a project.

Further collaboration can be used to establish the creation of new SOLES projects, ensuring that domains of most interest to researchers and funders are prioritised, and maximising crowd engagement to generate ideas for model, intervention and outcome assessment ontologies.

Furthermore, the ability of SOLES to provide a continually updated, comprehensive, annotated summary of current evidence, uncertainties and gaps of an entire research domain may prove instrumental in evidence-based policy making and research priority setting where broad-scope evidence synthesis produced within the timeframe of policy cycles are preferrable over traditional SRs which typically emphasises accuracy and depth of a narrower field [55].

Evolution of SOLES

SOLES projects are an emerging methodology, and we have thus far focussed our efforts on synthesising evidence from preclinical laboratory experiments within research-intensive areas where there is an urgent need to develop new therapeutics.

The evolution of the SOLES approaches began with a SR of animal models of depression which used machine learning for title and abstract eligibility screening and text-mining techniques to automatically extract key methodological details from publications including the type of animal model used and the pharmacological interventions tested [56]. The final categorised dataset of over 18,000 publications was then displayed in an interactive web application for others to interrogate.

Later, we expanded the SOLES approach to integrate more automation technologies, such as study quality and risk of bias annotation using a regular expressions approach, in an early version of our Alzheimer's disease dashboard. Further, by implementing a crowd-sourced approach to build up sets of annotated data at a quicker pace, we could train tailored automation approaches. Using this approach, we created a pilot COVID-SOLES, a resource which was updated on a weekly basis with new evidence in the emerging pandemic [57]. More recently we created the SPRINT-SOLES, in collaboration with over 20 experts in plant protection products who annotated subsets of the literature to use as training data for automated tools [58]. Harnessing this resource allowed researchers to produce rapid reviews of the evidence on the harms of pesticide exposures across clinical epidemiology, ecosystems and laboratory research. Our latest SOLES project, AD-SOLES [59], is the first example of a fully automated workflow from citation collection, screening for relevance, and organising data by risk of bias reporting, transparency measures, and study design characteristics in the field of Alzheimer's disease research.

In parallel, we have created MND-SOLES-CT [7,60,61] to inform selection and prioritisation of candidate drugs for evaluation in clinical trials in motor neuron disease (MND). MND-SOLES-CT reports current curated evidence from Repurposing Living Systematic Review – MND, a three-part machine learning assisted living systematic review of clinical studies of MND and other neurodegenerative diseases which may share similar pathways, MND animal *in vivo* studies and MND *in vitro* studies.

Bringing together the advances and learning points from each of these projects, we are now in the position to develop a functional automated workflow that can be replicated to form the skeleton of all new and ongoing SOLES projects. Our early SOLES approaches made great use of the R programming language and R Shiny Apps (RRID:SCR_001626). As the technological functionality has evolved, subsequent SOLES approaches have drawn in more programming tools such as database APIs, SQL databases and tools built in Python (RRID:SCR_008394), capitalising on interoperability.

Table 2 Exemplar SOLES project – AD-SOLES workflow

Stage	Details	Integrated tools
Systematic search	● Searches conducted using general Alzheimer's disease search terms across Web of Science, Scopus, EMBASE and PubMed ● Customised R packages which make use of APIs to retrieve data on a weekly basis for three databases, while EMBASE is searched manually every 6 months ● Additional metadata for each study retrieved from CrossRef and OpenAlex	rcrossref [46] openalexR [47] RISmed [48] ScopusAPI [49] rwos [50]
Automated deduplication	● When new citations are retrieved, a check is performed to see if they already exist in the SOLES database ● Automated deduplication tool identifies duplicates between new citations and existing citations and within new citations (across databases)	ASySD [39]
Machine-assisted screening	● Trained machine classifier (hosted at EPPI Centre, University College London) determines likelihood of a citation describing experiments in AD animal models	Machine classifiers [38]
Automated data annotation	● Full texts downloaded using open access APIs and publisher (dependent on institutional access) APIs ● Risk of bias reporting assessed using automated tool ● Open access status determined using CrossRef and OpenAlex APIs ● Open data and code status determined by X tool	rcrossref [46] pre-rob tool [41] openalexR [47] ODDPub [51]
Interactive shiny app	● Shiny application available at https://camarades.shinyapps.io/AD-SOLES/ ● This allows users to: 1. Gain an overview of the number of relevant publications in different categories, and intersections of those categories	
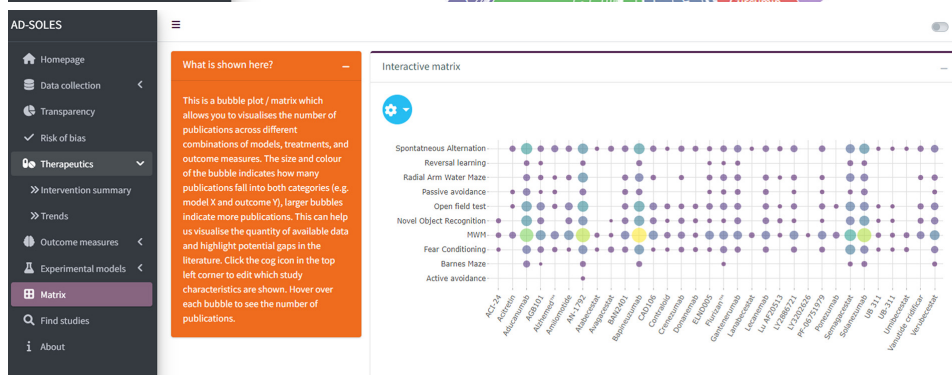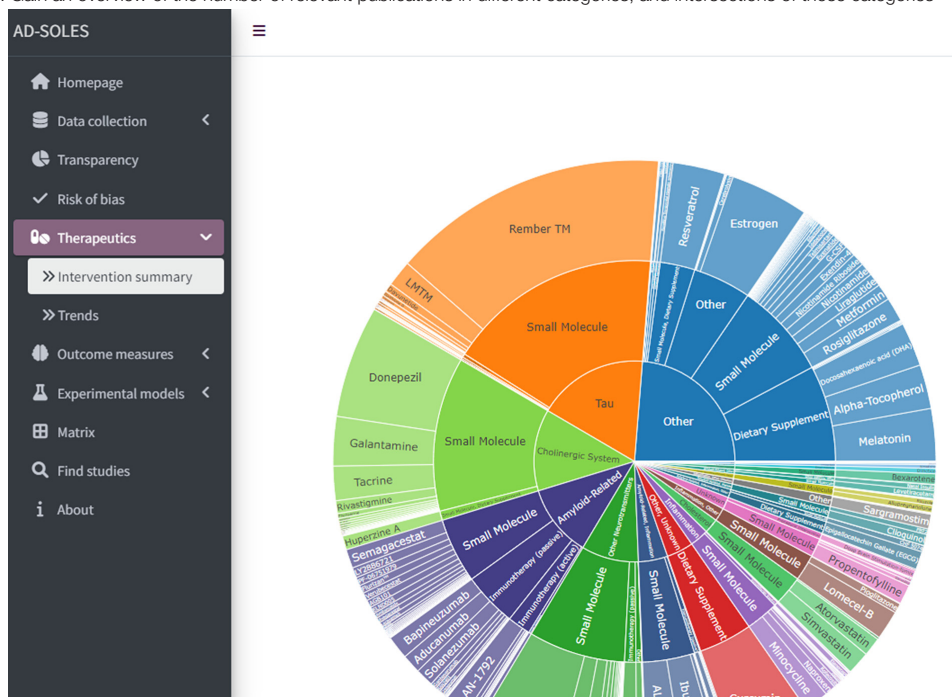
2. Visualise the proportion of publications with certain transparency and quality measures over time

Table 2 Exemplar SOLES project – AD-SOLES workflow (Continued)

Stage	Details	Integrated tools

3. Download a highly filtered citation lists to start a more detailed systematic literature reviews

Limitations of the SOLES approach

There are several limitations within the automated SOLES approaches we have outlined. Publications with no indexed abstract text may be incorrectly handled by machine learning-assisted screening approaches, while publications with no accessible full-text documents cannot be further categorised programmatically. The evolution of scientific publishing towards open-access publishing and improvements to the way citations are indexed in biomedical databases (e.g., to include abstracts with every publication in bibliographic databases) will support evidence synthesis approaches such as SOLES moving forward. The limitations of supervised machine learning algorithms for record screening of large corpuses have been discussed elsewhere (Bannach-Brown et al., 2019).

A further barrier to a fully automated approach, is that the way preclinical research publications are currently structured does not allow for accurate detection of PICO elements for every experiment. A publication may describe several experiments in different animal models, testing different treatments and measuring those animals on different outcomes. Further, it is difficult to determine how many regex matches indicate that a PICO element is part of an experiment described within a publication, rather than a match to a reference to other research.

These caveats suggest that while automated approaches are useful, they cannot completely replace human reviewers; a SOLES project will still require some human input to ensure quality of the data that is output.

Future development

Our key short-term goal for SOLES is to continually optimise the automated approaches we use, improving the performance of existing tools and optimising our approaches to take advantage of new tools which may become available. In particular, we recognise the growing need for high-level database management functionality for running automated scripts and updating records, and the implications this has in maintaining the quality of synthesised data.

In the long-term, we hope to focus on the interface between SOLES and SRs derived from SOLES projects. Interoperability between SOLES data visualisation apps and SR platforms such as the Systematic Review Facility (SyRF; [62]) strengthens the community and crowd approach to working collaboratively. This would benefit both SOLES developers and researchers conducting SRs, as additional data collected from derived SRs could feed back into a broader SOLES overview. For SRs derived from SOLES that share similar methodology, further efficiency gains could be made by developing an overarching master protocol with shared workflows and infrastructure for in-depth annotation, data extraction, analysis, reporting and visualisation. The SOLES approach and its flexibility means it is ideal to display outputs from other types of evidence synthesis, for example, network meta-analysis, systematic maps and scoping reviews.

Having developed SOLES across several research domains, we hope to expand our approach and triangulate diverse evidence streams including a greater focus on clinical evidence and results from *in vitro* experiments. Furthermore, the same interventions, drug targets and underlying biological pathways may be studied within different contexts. Combining evidence from across SOLES projects could enable us to uncover hidden links across typically distinct research silos. While aspects of SOLES overlap with evidence gap maps, we aim to leverage novel automation tools to support and expand our approach going forward to provide more advanced insights and drive hypothesis generation.

Finally, we seek greater collaboration with key research stakeholders who may benefit from the data that a SOLES can provide, such as laboratory researchers, clinicians and other healthcare decision makers. Only with effective collaboration can we ensure that SOLES are fit for purpose and beneficial to the wider scientific community.

Concluding remarks

Systematic Online Living Evidence Summaries are an emerging approach to benefit multiple key stakeholders in the evidence synthesis process, including primary researchers, evidence synthesists, information specialists and statisticians, funders, regulatory agencies and policymakers.

We believe this approach can bring these stakeholders together to work collaboratively to answer pressing research questions in biomedicine, forming communities to bridge the translational gap and make the most effective use of research.

The approach is flexible, data processing pipelines can be expanded to add new validated tools as they become available, to meet the diverse and growing needs of evidence stakeholders in biomedicine.

Data Availability

We did not generate nor analyse any new data for this manuscript.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

CRediT Author Contribution

Kaitlyn Hair: Conceptualization, Data curation, Software, Investigation, Visualization, Methodology, Writing—original draft, Writing—review & editing. **Emma Wilson:** Conceptualization, Visualization, Methodology, Writing—original draft, Writing—review & editing. **Charis Wong:** Conceptualization, Data curation, Software, Investigation, Visualization, Methodology, Writing—review & editing. **Anthony Tsang:** Methodology, Writing—review & editing. **Malcolm Macleod:** Conceptualization, Supervision, Funding acquisition, Methodology, Project administration, Writing—review & editing. **Alexandra Bannach-Brown:** Conceptualization, Data curation, Software, Investigation, Visualization, Methodology, Project administration, Writing—review & editing.

Acknowledgements

Abbreviations

MND, Motor Neuron Disease; PICO, Population, Intervention, Comparison and Outcomes; SOLES, Systematic Online Living Evidence Summaries.

References

1 Chalmers, I., Hedges, L.V. and Cooper, H. (2002) A brief history of research synthesis. *Eval. Health Profess.* **25**, 12–37, https://doi.org/10.1177/0163278702025001003

2 Mullen, P.D. and Ramírez, G. (2006) The promise and pitfalls of systematic reviews. *Annu. Rev. Public Health* **27**, 81–102, https://doi.org/10.1146/annurev.publhealth.27.021405.102239

3 van Luijk, J., Bakker, B., Rovers, M.M., Ritskes-Hoitinga, M., de Vries, R.B. and Leenaars, M. (2014) Systematic reviews of animal studies; missing link in translational research? *PLoS ONE* **9**, e89981, https://doi.org/10.1371/journal.pone.0089981

4 Hooijmans, C.R. and Ritskes-Hoitinga, M. (2013) Progress in using systematic reviews of animal studies to improve translational research. *PLoS Med.* **10**, e1001482, https://doi.org/10.1371/journal.pmed.1001482

5 Macleod, M.R., O'Collins, T., Howells, D.W. and Donnan, G.A. (2004) Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke* **35**, 1203–1208, https://doi.org/10.1161/01.STR.0000125719.25853.20

6 Sena, E.S., van der Worp, H.B., Bath, P.M.W., Howells, D.W. and Macleod, M.R. (2010) Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol.* **8**, e1000344, https://doi.org/10.1371/journal.pbio.1000344

7 Wong, C., Gregory, J.M., Liam, J., Egan, K., Vesterinen, H.M., Khan, A.A. et al. (2023) Systematic, comprehensive, evidence-based approach to identify neuroprotective interventions for motor neuron disease: using systematic reviews to inform expert consensus.. *BMJ Open* **13**, e064169, https://doi.org/10.1136/bmjopen-2022-064169

8 Wong, C., Stavrou, M., Elliott, E., Gregory, J.M., Leigh, N., Pinto, A.A. et al. (2021) Clinical trials in amyotrophic lateral sclerosis: a systematic review and perspective. *Brain Commun.* **3**, fcab242, https://doi.org/10.1093/braincomms/fcab242

9 Vesterinen, H.M., Connick, P., Irvine, C.M.J., Sena, E.S., Egan, K.J., Carmichael, G.G. et al. (2015) Drug repurposing: a systematic approach to evaluate candidate oral neuroprotective interventions for secondary progressive multiple sclerosis. *PLoS ONE* **10**, e0117705, https://doi.org/10.1371/journal.pone.0117705

10 Bornmann, L. and Mutz, R. (2015) Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *J. Assoc. Information Sci. Technol.* **66**, 2215–2222, https://doi.org/10.1002/asi.23329

11 Bashir, R., Surian, D. and Dunn, A.G. (2018) Time-to-update of systematic reviews relative to the availability of new evidence. *System. Rev.* **7**, 195, https://doi.org/10.1186/s13643-018-0856-9

12 Bastian, H., Glasziou, P. and Chalmers, I. (2010) Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med.* **7** (9), e1000326, https://doi.org/10.1371/journal.pmed.1000326

13 Créquit, P., Trinquart, L., Yavchitz, A. and Ravaud, P. (2016) Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer. *BMC Med.* **14**, 8, https://doi.org/10.1186/s12916-016-0555-0

14 Hair, K. (2022) Developing automated meta-research approaches in the preclinical Alzheimer's disease literature. University of Edinburgh, Edinburgh, Scotland, UK., Ph.D. Thesis

15 Wilson, E., Cruz, F., Maclean, D., Ghanawi, J., McCann Sarah, K., Brennan Paul, M. et al. (2023) Screening for in vitro systematic reviews: a comparison of screening methods and training of a machine learning classifier. *Clin. Sci.* **137**, 181–193, https://doi.org/10.1042/CS20220594

16 Tricco, A.C., Brehaut, J., Chen, M.H. and Moher, D. (2008) Following 411 Cochrane Protocols to completion: a retrospective cohort study. *PLoS ONE* **3**, e3684, https://doi.org/10.1371/journal.pone.0003684

17 Beller, E.M., Chen, J.K.-H., Wang, U.L.-H. and Glasziou, P.P. (2013) Are systematic reviews up-to-date at the time of publication? *Systematic Rev.* **2**, 36, https://doi.org/10.1186/2046-4053-2-36

18 Thomas, J., Noel-Storr, A., Marshall, I., Wallace, B., McDonald, S., Mavergames, C. et al. (2017) Living systematic reviews: 2. Combining human and machine effort. *J. Clin. Epidemiol.* **91**, 31–37, https://doi.org/10.1016/j.jclinepi.2017.08.011

19 Currie, , G.L., Angel-Scott, H.N., Colvin, L., Cramond, F. and Hair, K. (2019) Animal models of chemotherapy-induced peripheral neuropathy: A machine-assisted systematic review and meta-analysis.. *PLOS Biology* **17** (5), e3000243, https://doi.org/10.1371/journal.pbio.3000243

20 Nepomuceno, V. and Soares, S. (2019) On the need to update systematic literature reviews. *Information Software Technol.* **109**, 40–42, https://doi.org/10.1016/j.infsof.2019.01.005

21 Lefebvre, C., Glanville, J., Wieland, L.S., Coles, B. and Weightman, A.L. (2013) Methodological developments in searching for studies for systematic reviews: past, present and future? *Syst Rev.* **2**, 78, https://doi.org/10.1186/2046-4053-2-78

22 Shojania, K.G., Sampson, M., Ansari, M.T., Ji, J., Garritty, C., Rader, T. et al. (2007) *Updating Systematic Reviews*, Agency for Healthcare Research and Quality (US), Rockville (MD)

23 Garner, P., Hopewell, S., Chandler, J., MacLehose, H., Schünemann, H.J., Akl, E.A. et al. (2016) When and how to update systematic reviews: consensus and checklist. *BMJ* **354**, i3507, https://doi.org/10.1136/bmj.i3507

24 Shojania, K.G., Sampson, M., Ansari, M.T., Ji, J., Doucette, S. and Moher, D. (2007) How quickly do systematic reviews go out of date? A survival analysis. *Ann. Intern. Med.* **147** (4), 224–33, https://doi.org/10.7326/0003-4819-147-4-200708210-00179

25 Andersen, M.Z., Gülen, S., Fonnes, S., Andresen, K. and Rosenberg, J. (2020) Half of Cochrane reviews were published more than 2 years after the protocol. *J. Clin. Epidemiol.* **124**, 85–93, https://doi.org/10.1016/j.jclinepi.2020.05.011

26 Hoffmeyer, B.D., Andersen, M.Z., Fonnes, S. and Rosenberg, J. (2021) Most Cochrane reviews have not been updated for more than 5 years. *J. Evidence-Based Med.* **14**, 181–184, https://doi.org/10.1111/jebm.12447

27 Siontis, K.C., Hernandez-Boussard, T. and Ioannidis, J.P.A. (2013) Overlapping meta-analyses on the same topic: survey of published studies. *BMJ* **347**, f4501–f, https://doi.org/10.1136/bmj.f4501

28 Helfer, B., Prosser, A., Samara, M.T., Geddes, J.R., Cipriani, A., Davis, J.M. et al. (2015) Recent meta-analyses neglect previous systematic reviews and meta-analyses about the same topic: a systematic examination. *BMC Med.* **13**, 82, https://doi.org/10.1186/s12916-015-0317-4

29 Nakagawa, S., Dunn, A.G., Lagisz, M., Bannach-Brown, A., Grames, E.M. et al. (2020) A new ecosystem for evidence synthesis. *Nat. Ecol. Evolution* **4**, 498–501, https://doi.org/10.1038/s41559-020-1153-2

30 Colquhoun, H.L., Levac, D., O'Brien, K.K., Straus, S., Tricco, A.C., Perrier, L. et al. (2014) Scoping reviews: time for clarity in definition, methods, and reporting. *J. Clin. Epidemiol.* **67**, 1291–1294, https://doi.org/10.1016/j.jclinepi.2014.03.013

31 Saran, A. and White, H. (2018) Evidence and gap maps: a comparison of different approaches. *Campbell Systematic Rev.* **14**, 1–38, https://doi.org/10.4073/cmdp.2018.2

32 Miake-Lye, I.M., Hempel, S., Shanman, R. and Shekelle, P.G. (2016) What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Systematic Rev.* **5**, 28, https://doi.org/10.1186/s13643-016-0204-x

33 Bannach-Brown, A., Hair, K., Bahor, Z., Soliman, N., Macleod, M. and Liao, J. (2021) Technological advances in preclinical meta-research. *BMJ Open Sci.* **5**, e100131, https://doi.org/10.1136/bmjos-2020-100131

34 Marshall, I.J., Noel-Storr, A., Kuiper, J., Thomas, J. and Wallace, B.C. (2018) Machine learning for identifying Randomized Controlled Trials: An evaluation and practitioner's guide. *Res. Synthesis Methods* **9**, 602–614, https://doi.org/10.1002/jrsm.1287

35 Soto, A.J., Przybyla, P. and Ananiadou, S. (2018) Thalia: semantic search engine for biomedical abstracts. *Bioinformatics* **35**, 1799–1801, https://doi.org/10.1093/bioinformatics/bty871

36 O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M. and Ananiadou, S. (2015) Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Syst. Rev.* **4**, 5, https://doi.org/10.1186/2046-4053-4-5

37 Marshall, I.J. and Wallace, B.C. (2019) Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *System. Rev.* **8**, 163, s13643-019-1074-9, https://doi.org/10.1186/s13643-019-1074-9

38 Bannach-Brown, A., Przybyla, P., Thomas, J., Rice, A.S.C., Ananiadou, S., Liao, J. et al. (2019) Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. *System. Rev.* **8**, 23, https://doi.org/10.1186/s13643-019-0942-7

39 Hair, K., Bahor, Z., Macleod, M., Liao, J. and Sena, E.S. (2021) The Automated Systematic Search Deduplicator (ASySD): a rapid, open-source, interoperable tool to remove duplicate citations in biomedical systematic reviews. *bioRxiv*, https://doi.org/10.1101/2021.05.04.442412

40 Wang, , Q., Liao, J., Lapata, M. and Macleod, M. (2022) PICO entity extraction for preclinical animal literature. . *Syst Rev* **11**, 209, https://doi.org/10.1186/s13643-022-02074-4

41 Wang, Q., Liao, J., Lapata, M. and Macleod, M. (2022) Risk of bias assessment in preclinical literature using natural language processing. *Res. Synth. Meth.* **13**, 368–380, https://doi.org/10.1002/jrsm.1533

42 Elliott, J.H., Turner, T., Clavisi, O., Thomas, J., Higgins, J.P.T., Mavergames, C. et al. (2014) Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Med.* **11**, e1001603, https://doi.org/10.1371/journal.pmed.1001603

43 Bahor, Z., Liao, J., Macleod, M.R., Bannach-Brown, A., McCann, S.K., Wever, K.E. et al. (2017) Risk of bias reporting in the recent animal focal cerebral ischaemia literature. *Clin. Sci. (London, England : 1979)* **131**, 2525–2532, https://doi.org/10.1042/CS20160722

44 Al-Zubidy, A., Carver, J.C., Hale, D.P. and Hassler, E.E. (2017) Vision for SLR tooling infrastructure: prioritizing value-added requirements. *Information Software Technol.* **91**, 72–81, https://doi.org/10.1016/j.infsof.2017.06.007

45 van Altena, A.J., Spijker, R. and Olabarriaga, S.D. (2019) Usage of automation tools in systematic reviews. *Res. Synthesis Methods* **10**, 72–82, https://doi.org/10.1002/jrsm.1335

46 Chamberlain, , S., Zhu, H., Jahn, N., Boettiger, C. and Ram, K. rcrossref: Client for Various "CrossRef" "APIs" 2020. Available from: https://CRAN.R-project.org/package=rcrossref., Date of Access 16.01.2023

47 Le, M.A.T. (2023) openalexR: Getting Bibliographic Records from 'OpenAlex' Database Using 'DSL' API. Available from: https://cran.r-project.org/web/packages/openalexR/, Date of Access 16.01.2023

48 Kovalchik, S. RISmed: Download Content from NCBI Databases. https://cran.r-project.org/web/packages/RISmed/RISmed.pdf., Date of Access: 16.01.2023.

49 Belter, C. (2021) scopusAPI. R package version. https://github.com/christopherBelter/scopusAPI, Date of Access: 16.01.2023

50 Barnier, J. (2020) rwos: Interface to Web of Science Web Services API. https://rdrr.io/github/juba/rwos/

51 Riedel, N., Kip, M. and Bobrov, E. (2020) ODDPub – a Text-Mining Algorithm to Detect Data Sharing in Biomedical Publications.. *Data Science Journal* **19** (42), 1–14, https://doi.org/10.5334/dsj-2020-042

52 Menke, J., Eckmann, P., Ozyurt, I.B., Roelandse, M., Anderson, N., Grethe, J. et al. (2022) Establishing institutional scores with the rigor and transparency index: large-scale analysis of scientific reporting quality. *J. Med. Internet Res.* **24**, e37324, https://doi.org/10.2196/37324

53 Wang, Q., Hair, K., Macleod, M.R., Currie, G., Bahor, Z., Sena, E.S. et al. (2021) Protocol for an analysis of in vivo reporting standards by journal, institution and funder. preprint. *MetaArXiv Preprints* 2021

54 Cheng, J., Teevan, J., Iqbal, S.T. and Bernstein, M.S. (eds) (2015) Break it down: a comparison of macro- and microtasks. In *CHI '15: CHI Conference on Human Factors in Computing Systems*, ACM, Seoul Republic of Korea, 2015-04-18

55 Greenhalgh, T. and Malterud, K. (2017) Systematic reviews for policymaking: muddling through. *Am. J. Public Health* **107**, 97–99, https://doi.org/10.2105/AJPH.2016.303557

56 Bannach-Brown, A. (2018) Preclinical Models of Depression. https://camarades.shinyapps.io/Preclinical-models-of-depression/ [updated 2018]

57 Hair, K., Sena, E.S., Wilson, E., Currie, G., Macleod, M., Bahor, Z. et al. (2021) Building a systematic online living evidence summary of COVID-19 research. *J. EAHIL* **17**, 21–26, https://doi.org/10.32384/jeahil17465

58 Hair, K. (2022) SPRINT-SOLES. https://camarades.shinyapps.io/SPRINT-SOLES/ [updated 2022]

59 Hair, K. (2022) AD-SOLES. https://camarades.shinyapps.io/AD-SOLES/ [updated 2022]

60 Wong, C. (2022) MND-SOLES-CT_demo. [updated 2022] https://camarades.shinyapps.io/MND-SOLES-CT-demo/

61 Wong, C. (ed) (2022) *Developing a data-driven framework to identify evaluate and prioritise candidate drugs for motor neuron disease clinical trials*, ENCALS

62 Bahor, Z., Liao, J., Currie, G., Ayder, C., Macleod, M., McCann, S.K. et al. (2021) Development and uptake of an online systematic review platform: the early years of the CAMARADES Systematic Review Facility (SyRF). *BMJ Open Sci.* **5**, e100103, https://doi.org/10.1136/bmjos-2020-100103